working paper
department
of economics

SPECIFICATION TESTS IN ECONOMETRICS

J. A. Hausman

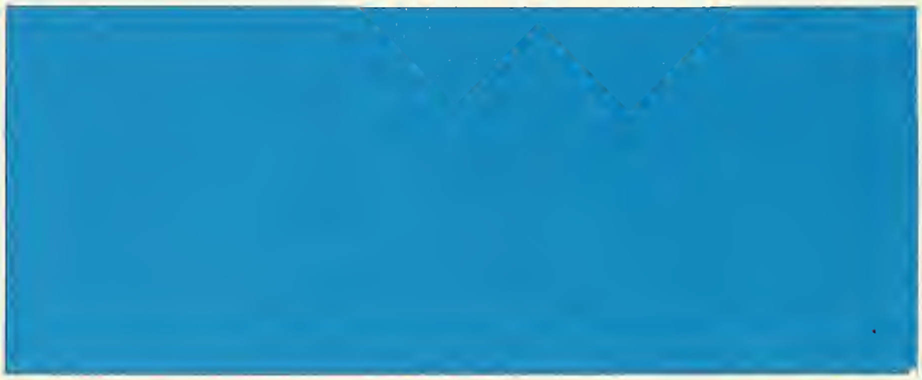
MASS. INST. TECH
OCT 13 1976
DEWEY LIBRARY

Number 185

June 1976
Revised August 1976

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139



SPECIFICATION TESTS IN ECONOMETRICS

J. A. Hausman

MASS. INST. TECH
OCT 13 1976
DEWEY LIBRARY

Number 185

June 1976
Revised August 1976

I would like to thank D. W. Carlton, G. Chamberlain, F. M. Fisher, R. H. Gordon, R. E. Hall, and H. L. White for helpful discussions. A. S. Kelso and E. R. Rosenthal provided research assistance. Research support has been provided by the NSF.

The views expressed in this paper are the author's sole responsibility and do not reflect those of the Department of Economics, the Massachusetts Institute of Technology, or the National Science Foundation.

Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

http://www.archive.org/details/specificationtes00haus

1. Introduction

Specification tests form one of the most important areas for research in econometrics. In the standard regression framework, $y = X\beta + \varepsilon$, the two stochastic specifications are first that ε is independent of X (or for fixed X, ε has expectation zero) and that ε has a spherical covariance matrix

(1.1) a. $E(\varepsilon|X) = 0$

b. $V(\varepsilon|X) = \sigma^2 I$

Failure of the first assumption, sometimes called the orthogonality assumption, leads to biased estimates while failure of the second assumptions, sometimes called the sphericality assumption, leads to loss of efficiency although the central tendency of the estimator is still correct. While in many problems the payoff to detecting failure of assumption (1.1.a.) is presumably greater than detecting failure of assumption (1.1.b.), most of the attention in the econometric literature has been paid to devising tests for the latter assumption. Theil's [1957] famous specification test for left out variables is almost the only exception. Yet, the problem is so important that increased attention should be paid, especially since efficient estimators are now used in almost all situations; and these estimators are often quite sensitive to failures of the first assumption.

In this paper a general form of specification test is proposed which attempts to provide powerful tests of assumption (1.1.a.). A main stumbling block to specification tests has been a lack of precisely specified alternative hypotheses. Here, I point out that in many situations, including time series - cross section specifications, errors in variables

0723954

specifications, and simultaneous equation specifications, the alternative hypothesis that assumption (1.1.a.) fails may be tested in an expanded regression framework. The basic idea follows from the existence of an alternative estimator which is unbiased under both null and alternative hypotheses. By comparing the estimates from this estimator with the efficient estimator (under assumption 1.1.a.) and noting that their difference is uncorrelated with the efficient estimator when the null hypothesis is true, easily used tests may be devised which have the form

$$(1.2) \quad y = X\beta + \tilde{X}\alpha + v$$

where $\tilde{X}$ is a suitably transformed version of X. These tests are easily performed using standard regression programs to test H_0: $\alpha = 0$. Furthermore, the power of the test may be ascertained by considering the estimated variance of $\hat{\alpha}$. Power considerations are important when the null hypothesis is not rejected to determine how much evidence is present that a Type II error is not being committed.

In Section 2 the basic lemma regarding these types of specification tests is proven. The test is applied to an errors in variables problem and equation (1.2) is derived. In the next section the relationship to Theil's result is indicated. The following two sections discuss two new specification tests for the time series - cross section model and for the simultaneous equation model. Both tests are always available (unlike the errors in variables test which requires an instrumental variable) and should be used for these two important model specifications. In Section 6 the issues of pretesting and minimum mean square error estimation are discussed within the context of specification error tests. Lastly,

two examples are provided. The first example is especially interesting since a widely used time series - cross section specification, the random effects model is decisively rejected. Considerable doubt is thereby cast on much cross section analysis of individual data. The general principle of this paper can be applied in additional problems not considered here. Therefore the tests should be useful to the applied econometrician.

2. Theory and a Test of Errors in Variables

The theory underlying the proposed specification tests rests on one fundamental idea. Under the (null) hypothesis of no misspecification, there will exist an unbiased and efficient estimator, where efficiency means attaining the Cramer-Rao bound.[1] Under the alternative hypothesis of misspecification, however, this estimator will be biased. To construct a test of misspecification, it is necessary to find another estimator which is not adversely affected by the misspecification; but this estimator will not be efficient under the correct specification. A consideration of the difference between the two estimates, $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$ where $\hat{\beta}_0$ is the efficient estimate under H_0 and $\hat{\beta}_1$ is an appropriate estimator under H_1, will then lead to a specification test. If no misspecification is present, the expected value of $\hat{q}$ is zero. With misspecification $\hat{q}$ will differ from zero; and if the power of the test is high, $\hat{q}$ will be large relative to its standard error. Hopefully, this procedure will lead to powerful tests in important cases because the misspecification is apt to be serious only when the two estimates differ substantially.

In constructing tests based on $\hat{q}$, an immediate problem comes to mind. To develop tests not only is the expected value of $\hat{q}$ required, but the variance of $\hat{q}$, $V(\hat{q})$, must also be determined. Since $\hat{\beta}_0$ and $\hat{\beta}_1$ use the same data, they will be correlated which could lead to a messy calculation for the variance of $\hat{q}$. Luckily, this problem is resolved easily and, in fact, $V(\hat{q}) = V(\hat{\beta}_1) - V(\hat{\beta}_0)$ under the null hypothesis of no misspecification.

---

1. Since the goal is to develop tests, a normal distribution is assumed throughout for the disturbances. Ordinary least squares may also be thought to be 'efficient' in the sense of being the Gauss-Markov estimator. For large sample estimators, the properties of consistency and asymptotic efficiency are relevant.

Thus, the construction of specification error tests is simplified since the estimators may be considered separately without regard to their interaction. The intuitive reasoning behind this result is simple although it appears to have remained unnoticed in the general case. The idea rests on the fact that the efficient estimator, $\hat{\beta}_0$, must be uncorrelated with $\hat{q}$ under the null hypothesis for any other unbiased estimator $\hat{\beta}_1$. If this were not the case, a linear combination of $\hat{\beta}_0$ and $\hat{q}$ could be taken which would lead to an unbiased estimator $\hat{\beta}_*$ which would have smaller variance than $\hat{\beta}_0$ which is assumed efficient. To prove the result formally, consider the following lemma:

(2.1) Lemma: Consider two estimators $\hat{\beta}_0$, $\hat{\beta}_1$ which are both unbiased and normally distributed with $\hat{\beta}_0$ attaining the Cramer-Rao bound (alternatively, both consistent and asymptotically normal with $\hat{\beta}_0$ attaining the CR bound asymptotically[1].). Consider $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$. Then $\hat{\beta}_0$ and $\hat{q}$ have zero covariance, $C(\hat{\beta}_0, \hat{q}) = 0$.

Proof: Suppose $\hat{\beta}_0$ and $\hat{q}$ are not orthogonal. Since $E\hat{q} = 0$ define a new estimator $\hat{\beta}_2 = \hat{\beta}_0 + rA\hat{q}$ where r is a scalar and A is an arbitrary matrix to be chosen. The new estimator is unbiased and normal with variance

$$(2.2)\quad V(\hat{\beta}_2) = V(\hat{\beta}_0) + rAC(\hat{\beta}_0, \hat{q}) + rC(\hat{\beta}_0, \hat{q})A' + r^2AV(\hat{q})A'.$$

Now consider the difference between the variance of the new estimator and the old efficient estimator

1. Besides consistency and asymptotic normality, uniform convergence is also required to rule out superefficiency. See Rao [1973, p. 284].

$$(2.3)\quad F(r) = V(\beta_2) - V(\beta_0) = rAC + rCA' + r^2 AV(\hat{q})A'$$

Taking derivatives with respect to r yields

$$(2.4)\quad F'(r) = AC + CA' + 2rAV(\hat{q})A'.$$

Then choose $A = -C'$ and noting that C is symmetric leads to

$$(2.5)\quad F'(r) = -2C'C + 2rC'V(\hat{q})C.$$

Therefore at $r = 0$, $F'(0) = -2C'C \leq 0$ in the sense of being nonpositive definite. But $F(0) = 0$ so for r small there is a contradiction unless $C = C(\hat{\beta}_0, \hat{q}) = 0$ since $\hat{\beta}_0$ is efficient.

Once it has been shown that the efficient estimator is uncorrelated with $\hat{q}$, the variance of $\hat{q}$ is easily calculated.

(2.6) <u>Corollary</u>: $V(\hat{q}) = V(\hat{\beta}_1) - V(\hat{\beta}_0) \geq 0$ in the sense of being nonnegative definite.

Proof: Since $\hat{q} + \hat{\beta}_0 = \hat{\beta}_1$, $V(\hat{q}) + V(\hat{\beta}_0) = V(\hat{\beta}_1)$. Furthermore, $\hat{\beta}_0$ attains the CR bound.

Given this result a general misspecification test can be specified by considering the statistic

$$(2.7)\quad m = \hat{q}'V(\hat{q})^{-1}\hat{q} \sim KF_{K,T-K}$$

where K is the number of unknown parameters in β, and F is distributed as Snedecor's central F distribution with K and T-K degrees of freedom when no misspecification is present.[1.]

1. In forming $V(\hat{q})$ the estimate of σ^2 used must be independent of $\hat{q}$ for m to be distributed exactly as F. To insure this property and also for the analysis of the noncentral F considered below the estimate of σ^2 from $\hat{\beta}_1$, s_1^2, is used. For the case where other elements of $V(\hat{q})$ are estimated, e.g., the simultaneous equation problem of section 5, then large sample properties are used and m is distributed as χ^2_K.

The statistic m in equation (2.7) specifies the distribution of the difference of the two estimators when no misspecification is present. The other operating characteristic of a test is its power. Unfortunately, power considerations have not been paid much attention in econometrics probably due to the impreciseness of alternative hypotheses. The power of the statistic in equation (2.7) depends on the noncentral F distribution with noncentrality parameter δ^2

$$(2.8) \quad \delta^2 = \bar{q}'V(\hat{q})^{-1}\bar{q}$$

where $\bar{q} = E(\hat{\beta}_1 - \hat{\beta}_0)$ the expected difference between the two estimates. For a given size of test the power increases with δ^2 which in turn depends on how far the biased estimator $\hat{\beta}_0$ is from the unbiased (consistent) estimator $\hat{\beta}_1$ when misspecification is present. Thus, the comparison estimator $\hat{\beta}_1$ should be chosen so that if a certain type of misspecification is feared to be present, $\hat{q}$ which is the difference of the estimates, is expected to be large. The other consideration in equation (2.8) is to keep $V(\hat{q})$ small so that a large departure between $\hat{\beta}_0$ and $\hat{\beta}$ will not arise by chance. This requirement means that $\hat{\beta}_1$ should be relatively efficient but at the same time sensitive to departures from the model specification. To highlight the power considerations the specification test of equation (2.7) will be reformulated in a statistically equivalent form which will keep these power considerations uppermost in the user's mind. Also, the reformulated test will be normally easier to use with available econometrics computer programs. To demonstrate this reformulated test, an errors in variables example is considered.

An errors in variables test attempts to determine if stochastic regressors and the disturbances in a regression are independent. In the simplest case consider the model

$$(2.9) \quad y_i = \beta x_i + \varepsilon_{1i} \qquad i = 1,\ldots,T$$

where y_i, x_i, and ε_i are all iid with mean zero and distributed normally. Under the null hypothesis, x_i and η_i are independent

$$(2.10) \quad C(x_i, \varepsilon_{1i}) = 0 \quad \text{for all } i \quad (\text{also plim } \tfrac{1}{T} x'\varepsilon_1 = 0)$$

while under the alternative hypothesis the covariance is nonzero.

The efficient estimator under the null hypothesis is, of course, least squares. Under the alternative hypothesis least squares is biased with H_1: $E\hat{\beta}_0 = \beta \dfrac{\sigma_x^2 - \sigma^2_2}{\sigma_x^2}$ where the observed $x_i = x_i^* + \varepsilon_{2i}$, the sum of the "true" regressor and a normal random variable with mean zero which is assumed independent of ε_{1i}. The comparison estimator $\hat{\beta}_1$ will be an instrumental variable (IV) estimator based on the instrument z with properties

$$(2.11) \quad C(z_i, \eta_i) = 0 \quad C(z_i, x_i) \neq 0 \quad (\text{or plim } \tfrac{1}{T} z'\eta = 0$$

$$\text{plim } \tfrac{1}{T} z'x \neq 0) \quad \text{for } \eta_i = \varepsilon_{1i} - \beta\varepsilon_{2i}$$

Then the IV estimator is

$$(2.12) \quad \hat{\beta}_1 = (z'x)^{-1} z'y$$

To form the test statistic conditional on x under the null hypothesis using corollary (2.6)

$$(2.13)\quad \hat{q} = \hat{\beta}_1 - \hat{\beta}_0 \sim N(0, B)$$

where $B = V(q) = \sigma^2[(\hat{x}'\hat{x})^{-1} - (x'x)^{-1}]$ where $\hat{x} = z(z'z)^{-1}z'x$. Again using the corollary $\hat{q}'B^{-1}\hat{q}$ is distributed as χ^2_1. To derive an F test use s_1^2 the IV estimator of σ^2 to form $\hat{B}$. Then the test of misspecification is

$$(2.14)\quad m = \hat{q}'\hat{B}^{-1}\hat{q} \sim F_{1,T-1}$$

Choice of an alternative estimate of σ^2, say s_0^2 the OLS estimate of σ^2, leads to a similar test which is distributed as χ^2_1 in large samples under the null hypothesis.[1.] The large sample approximation to the power of the test depends primarily on the numerator of equation (2.14) as shown in equation (2.8). Under H_1, the expected value of q, $\bar{q} = \beta\cdot\sigma^2_{\varepsilon_2}/\sigma^2_x$ so the power depends on the magnitude of the two coefficients and the correlation of the right hand side variable with the disturbance. To compute the power as a function of β, equation (2.8) can be used. The IV estimates, $\hat{\beta}_{IV}$ and s_1^2, are unbiased under both the null and alternative hypotheses as is $V(\hat{q})$. An unbiased estimate of σ^2_x follows from the data using the formula for the sample variance, and an estimate of $\sigma^2_{\varepsilon_2}$ is derived from the equation $\hat{\sigma}^2_{\varepsilon_2} = (1 - \hat{\beta}_{OLS}/\hat{\beta}_{IV})\hat{\sigma}^2_x$. Then $\bar{q}$ may be calculated for any choice of β and the noncentrality parameter δ^2 is a quadratic function around $\beta = 0$, $\delta^2 = (\beta^2\sigma^4_{\varepsilon_2}/\sigma^4_x V(\hat{q}))$. The tables of the noncentral F test in Scheffé [1959] can be consulted to find the probability of the null hypothesis being rejected for a given value of β if the alternative

---

1. Under the alternative hypothesis, the power of the test is difficult to analyze since $\hat{B}$ is now an inconsistent estimate of B.

hypothesis is true conditional on the estimates of the incidental parameters of the problem. This type of IV (instrumental variable) tests for errors in variables was first proposed by Liviatan [1963]. Wu [1973] generalizes the test and considers tests with different estimates of the nuisance parameter σ^2. He also calculates the power of the different tests.[1.]

The IV test for errors in variables is known in the literature, but an alternative formulation of test leads to easier implementation. Also, the alternative formulation demonstrates the general format of misspecification tests. Consider the regression specification with 2 arbitrary scalar parameters $\tilde{\beta}$ and α

$$(2.15) \quad y = x\tilde{\beta} + \hat{x}\alpha + v.$$

where as before $\hat{x} = z(z'z)^{-1}z'x = P_z x$. Define $Q_x = I - P_x$ where $P_x = x(x'x)^{-1}x'$ and project the model of equation (2.15) into the subspace orthogonal to x

$$(2.16) \quad Q_x y = Q_x x\hat{\beta} + Q_x \hat{x}\alpha + Q_x v$$

Taking expectations of equation (2.16) under the null hypothesis where from equation (2.9), $EQ_x y = y - x\beta$ and $EQ_x x\tilde{\beta} = 0$,

$$(2.17) \quad Ey = x\beta + Q_x \hat{x}\alpha.$$

---

1. The instrumental variable test can also be considered a formalization and an improvement of a suggestion by Sargan [ ] who recommended checking whether the least squares estimates lie outside the confidence regions of the IV estimates. For individual coefficients the procedure used here is to see whether the least squares estimate lies outside the confidence regions centered at the IV estimate and with length formed from the square root of the difference of the IV variance minus the OLS variance. Thus shorter confidence intervals follow from the current procedure than from Sargan's suggestion. The F test on all the coefficients in equation (2.14), however, is the preferred test of the null hypothesis rather than separate consideration of each confidence interval.

so that the expectation of the second term should be zero if the null hypothesis is true. Then estimating α from using OLS on equation (2.15) leads to an estimate

$$\text{(2.18)} \qquad \hat{\alpha}_0 = (\hat{x}'Q_x\hat{x})^{-1}\hat{x}'Q_x y.$$

A test of $\alpha = 0$ from equation (2.17) under the null hypothesis is then based on the statistic $\sigma^2\chi^2 = \hat{\alpha}_0'(\hat{x}'Q_x\hat{x})\hat{\alpha}_0$. But $\frac{1}{\sigma^2}(\hat{x}'Q_x\hat{x})^{-1} = (\hat{x}'\hat{x})^{-1}B^{-1}(\hat{x}'\hat{x})^{-1}$ and $\hat{\alpha}_0 = (\hat{x}'Q_x\hat{x})^{-1}(\hat{x}'\hat{x})\hat{q}$. Thus, this formulation is equivalent to the IV test of equation (2.14) since

$$\text{(2.19)} \quad \frac{1}{\sigma^2}\hat{\alpha}_0'(\hat{x}'Q_x\hat{x})\hat{\alpha}_0 = \frac{1}{\sigma^2}\hat{q}'(\hat{x}'\hat{x})(\hat{x}'Q_x\hat{x})^{-1}(\hat{x}'\hat{x})\hat{q}$$
$$= \hat{q}'B^{-1}\hat{q}.$$

A simple t-test on $\hat{\alpha}$ on the OLS estimate $\hat{\alpha}_0$ from equation (2.15) yields a test on whether errors in variables is present and is equivalent to the large sample test using s_0^2 under the null hypothesis since equation (2.17) shows that α equals zero under the null hypothesis of no errors in variables. Besides ease of computation another advantage is present. Three outcomes of the test will be encountered leading to simple power interpretations which may not be as evident using the previous formulation of the test. First, $\hat{\alpha}_0$ may be large relative to its standard error. This result points to rejection of the hypothesis of no misspecification. The other clear cut case is a small $\hat{\alpha}_0$ with a small standard error which presents little evidence against H_0. The last result is a large standard error relative to the size of $\hat{\alpha}_0$. This finding points to lack of power of the test and arises when x and $\hat{x}$ in equation (2.15) are "multicollinear" leading to a small $(\hat{x}'Q_x\hat{x})$. If z is not a very good instrument because it

is not highly correlated with x, then the estimated standard error will be large relative to $\hat{\alpha}_0$. The lack of power will be very evident to the user since he will not have a precise estimate of α.

Two immediate generalizations of the errors in variables specification test can be made. The test can be used to test any potential failure of Assumption (1.1.a.) that the right hand side variables are orthogonal to the error term so long as instrumental variables are available. First, additional right hand side variables can be present

$$(2.20) \quad y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

where the X_1 variables are possibly correlated with ε while the X_2 variables are known to be uncorrelated. Given a matrix of variables Z (which should include X_2), $\hat{q}$ will again be the difference between the IV estimator and the efficient OLS estimator. Letting $\hat{X}_1 = P_Z X_1$ leads to the regression

$$(2.21) \quad y = X_1\beta_1 + X_2\beta_2 + \hat{X}_1\alpha + v$$

where a test of H_0: $\alpha = 0$ is a test for errors in variables. The last orthogonality test involves a lagged endogenous variable which may be correlated with the disturbance. In this case, however, if the specification of the error process is known such as first order serial correlation, a more powerful test may be available.[1.]

---

1. For the true regression problem (no lagged endogenous variables) under both the null hypothesis of no serial correlation and the alternative hypothesis $\hat{\beta}_0$, the OLS estimator, is unbiased since only Assumption 1.1.b. is violated. Therefore, if the null hypothesis of serial correlation is tested with an autoregressive estimator $\hat{\beta}_1$, $E\hat{q} = \bar{q} = 0$ under <u>both</u> hypotheses. If $\hat{q}$ is large relative to its standard error, misspecification is likely to be present.

In this section the general nature of the misspecification problem has been discussed when there exists an alternative estimator which provides consistent estimates under misspecification. By demonstrating that the efficient estimator is uncorrelated with the difference between the consistent and efficient estimator, a simple expression for the variance of the test is found. Then by applying it to the errors in variables problem, a very easy method to apply the test is demonstrated which also makes power considerations clearer. Before going on to discuss additional specification tests, "the" original specification test of Theil is discussed, and the current approach is shown to be a generalization of Theil's analysis.

3. The Classic Misspecification Result

Theil's [1957] classic misspecification theorem concerns the bias introduced due to left out variables in a regression specification. While the result has been widely used to assess bias when variables are not available [e.g., Griliches [1957]], perhaps the techniques used in this paper will be clearer when their relationship is shown to Theil's seminal work.[1.] The true underlying model is

$$(3.1) \quad y = X\beta + Z\alpha + \varepsilon$$

and the analysis determines the effect on the OLS estimator $\hat{\beta}_0$ when Z is omitted so that the specification $y = X\beta + \varepsilon$ is used. Let the null hypothesis H_0 be that $\alpha = 0$ or that X and Z are orthogonal, while under the alternative hypothesis $\alpha \neq 0$ and X and Z are not orthogonal. Then the efficient estimator under H_0 is $\hat{\beta}_0 = (X'X)^{-1}X'y$ while the alternative estimator which is inefficient under H_0 but unbiased under H_1 is $\hat{\beta}_1 = (X'Q_ZX)^{-1}X'Q_Zy$ where $Q_Z = I-P_Z$ with $P_Z = Z(Z'Z)^{-1}Z'$. Thus, the difference between the two estimators is $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$ which can be used to construct the appropriate test. As the lemma guarantees, $\hat{q}$ and $\hat{\beta}_0$ are orthogonal, for checking their covariance leads to

$$(3.2) \quad C(\hat{q}, \hat{\beta}_0) = E[\hat{q}(\hat{\beta}_0-\beta)'] = E[(X'Q_ZX)^{-1}X'Q_Z - (X'X)^{-1}X'][X\beta + \varepsilon].$$
$$[\varepsilon'X(X'X)^{-1}] = 0$$

Thus the regression specification equivalent to equation (2.15) of the preceding section is

$$(3.3) \quad y = X\beta + Q_ZX\alpha + v.$$

---

1. The earliest reference to this result that I know is Bancroft [1944, p. 198] who derives the result when both X and Z are one variable vectors.

The test of $\alpha = 0$ from this regression is then equivalent to testing whether bias will be introduced due to Theil's misspecification theorem.

This misspecification test is often difficult to apply because observations on Z are not available either because data has not been collected or because Z is an unobservable variable. The next misspecification test, however, always can be done since the necessary data is available. It is a test on the random effects model which has been widely used in econometrics.

4. Time Series - Cross Section Models

Time series - cross section models have become increasing important in econometrics. Many surveys, rather than being limited to a single cross section, now follow a panel of individuals over time. These surveys lead to a rich body of data given the wide variability between individuals coupled with much less variability for a given individual over time. Another important use of these models is to estimate demand across states over a period of time. Since for many goods (e.g., energy) considerable price variation exists across states while aggregate price indices move smoothly over time, time series - cross section models allow disentanglement of income and substitution effects which is often difficult to do with aggregate data.

The simplest time series - cross section model is specified as

$$(4.1)\quad y_{it} = X_{it}\beta + \mu_i + \varepsilon_{it} \qquad i = 1,N;\ t = 1,T$$

where μ_i is the individual effect. The two alternative specifications of the model differ in their treatment of the individual effect. The so-called fixed effects model treats μ_i as a fixed but unknown constant differing across individuals. Therefore, least squares on equation (4.1) is the correct estimator. To estimate the slope coefficients, deviation from means are used leading to the transformed observations $\tilde{y}_{it} = y_{it} - \bar{y}_{i.}$, $\tilde{X}_{it} = X_{it} - \bar{X}_{i.}$, $\tilde{\varepsilon}_{it} = \varepsilon_{it} - E\bar{\varepsilon}_{i.}$ and the regression specification[1].

$$(4.2)\quad \tilde{y}_{it} = \tilde{X}_{it}\beta + \tilde{\varepsilon}_{it}.$$

1. Analysis of variance notation is being used, e.g., $\bar{y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$.

An equivalent way of writing equation (4.2) is to let e be a T column vector of ones so that $e = (1,1,\ldots,1)'$ and to let $P_e = e(e'e)^{-1}e' = \frac{1}{T}ee' = \frac{1}{T}J_T$ with $Q_e = I - P_e$. Then the fixed effects specification on the stacked model is

$$(4.3)\quad Q_e y = Q_e X\beta + Q_e\alpha + Q_e\varepsilon = \tilde{X}\beta + \tilde{\varepsilon}$$

which is identical to equation (4.2) since $Q_e\alpha = 0$.

The alternative specification for the time series - cross section model is known as the random effects or variance components model. Instead of treating μ_i as a fixed constant, this specification assumes that μ_i is drawn from an iid distribution, $\mu_i \sim N(0, \sigma_\mu^2)$, and is uncorrelated with the ε_i. The specification then becomes

$$(4.4)\quad y_{it} = X_{it}\beta + \eta_{it},\ \eta_{it} = \mu_i + \varepsilon_{it}$$

so that $E\eta = 0$ and the covariance matrix is block diagonal.

$$(4.5)\quad \Omega = V(\eta) = \begin{bmatrix} \sigma_\mu^2 J_T + \sigma_\varepsilon^2 I_T & & & 0 \\ & \ddots & & \\ & & \sigma_\mu^2 J_T + \sigma_\varepsilon^2 I_T & \\ & & \ddots & \\ 0 & & & \sigma_\mu^2 J_T + \sigma_\varepsilon^2 I_T \end{bmatrix}$$

Here the appropriate estimator is generalized least squares $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ which can be expressed in weighted least squares form by transforming the variables by $\check{y}_{it} = y_{it} - \gamma\bar{y}_{i.}$, $\check{X}_{it} = X_{it} - \gamma\bar{X}_{i.}$ and then running ordinary least squares where

$$(4.6)\quad \gamma = 1 - \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2}\right)^{\frac{1}{2}}.$$

Usually the variances, σ_μ^2 and σ_ε^2 are not known, so consistent estimates are derived from intial least squares estimates to form $\hat{\gamma}$ (see Wallace and Hussain [1969]). This estimator is asymptotically efficient; and, if iterated to convergence, it yields the maximum likelihood estimates.

The choice of specification seems to rest on two considerations, one logical and the other statistical. The logical consideration is whether the μ_i can be considered random and drawn from an iid distribution. Both Scheffé [1959] and Searle [1971] contain excellent discussion of this question within an analysis of variance framework. Another way to consider the problem, suggested by Gary Chamberlain, is to decide whether the μ_i's satisfy di Finnetti's exchangeability criterion which is both necessary and sufficient for random sampling. Briefly, the criterion is to consider the sample $\underset{\sim}{\mu} = (\mu_1,\ldots,\mu_N)$ and to see whether we can exchange μ_i and μ_j (e.g., the constant for Rhode Island and California) while maintaining our subjective distribution to be the same. If this logical criterion is satisfied, as it might well be for models of individuals like an earnings function, then the statistical consideration is to compare the bias and efficiency of the two specifications in estimating β, the slope coefficients. Wallace and Hussain [1969], Maddala [1971], and Nerlove [1971] have recently discussed this issue, all pointing out that the specifications become identical as T becomes large in the appropriate way as can be seen by the definition of γ in equation (4.6). Since the case in econometrics is usually that N is large relative to T, differences between the two estimators are an important problem.

Under the random effects specification (say with known Ω for simplicity), $\hat{\beta}_{GLS}$ is the Gauss-Markov estimator while the fixed effects estimator $\hat{\beta}_{FE}$

is unbiased but not efficient.[1] However, an important issue of specification arises which was pointed out by Maddala [1971, p. 357] and has been further emphasized by Mundlak [1976]. The specification issue is whether the μ_i can be regarded as independent of the X_{it}'s, i.e., whether $E(\mu_i | X_{it}) = 0$. If this assumption is violated, the random effects estimator is biased while the fixed effects specification remains unbiased since it orthogonalizes the specification with respect to the individual constants. Consider an individual earnings equation over time. If an unobserved variable, "spunk", affects education and has an additional effect on earnings, then the assumption of independent μ_i's will be violated. Thus, a natural test of the null hypothesis of independent μ_i's is to consider the difference between the two estimators, $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}$. If no misspecification is present, then $\hat{q}$ should be near zero. Using the lemma, $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{GLS})$ so a specification test follows from $m = \hat{q}'V(\hat{q})^{-1}\hat{q}$ where $V(\hat{q}) = (X'Q_eX)^{-1} - (X'\Omega^{-1}X)^{-1}$. If the random effects specification is correct the two estimates should be near each other, rather than differing widely as has been reported sometimes in the literature as a virtue of the random effects specification. Therefore, while Maddala [1971, p. 343] demonstrates that $\hat{\beta}_{GLS}$ is a matrix weighted average of $\hat{\beta}_{FE}$ (the within group estimator) and the between group estimator, if the specification is correct then $E\hat{q} = 0$ so $\hat{\beta}_{GLS}$ and $\hat{\beta}_{FE}$ should be almost the same within sampling error. When the econometrician finds his estimates $\hat{\beta}_{FE}$ to be unsatisfactory, this evidence is a finding against his specification, not his choice of estimator.

---

1. The problem again arises that with Ω unknown, equation (2.8) cannot be used to numerically calculate the power since under the alternative hypothesis $V(\hat{q})$ is an inconsistent estimator because $\hat{\Omega}$ is also an inconsistent estimator of Ω.

The equivalent test in the regression format is to test $\alpha = 0$ from doing least squares on

(4.7) $\check{y} = \check{X}\hat{\beta} + \tilde{X}\alpha + v.$

where $\check{y}$ and $\check{X}$ are the γ transformed random effects variables while $\tilde{X}$ are the deviations from means variables from the fixed effects specification. The tests can be shown to be equivalent using the methods of the previous two sections and the fact that $Q_e\check{y} = Q_e y$. This test is easy to perform since $\check{X}$ and $\tilde{X}$ differ only in the choice of γ from equation (4.6) while $\tilde{X}$ has $\gamma = 1$.

The regression specification of equation (4.7) again makes power considerations evident. The noncentrality parameter of the F-test is proportional to the correlation of X and α which is the null hypothesis being tested. If γ is near unity, then the two estimators will give similar results and $\hat{q}$ will be near zero. The test of α from equation (4.7) will depend on $\hat{q}$ and also on how close X and $\check{X}$ and $\tilde{X}$ are. If they are quite different, $V(\hat{q})^{-1}$ will be small and then $\hat{\alpha}$ will be precisely estimated. When they are similar, the specification test will not have much power, but this case is not so important since the two estimates of β will also be similar.

It will often be the case in econometrics that γ will not be near unity. In many applications σ_μ^2 is small relative to σ_ε^2; and the problem sometimes arises that when σ_μ^2 is estimated from the data it turns out to be negative. For a panel followed over time the X_{it} are often nearly constant or trend smoothly with time so that much of the interindividual variation disappears into the individual constants when the fixed effects

estimator is used. However, it seems preferable to have unbiased estimates of the remaining slope coefficients by using a fixed effects specification and then attempt to sort out the effects of education, "spunk", and their interaction through a parametrization of the individual constants. The misspecification test from equation (4.7) thus seems a desirable test of the two different specifications.

In this section a test of the implicit assumption behind the random effects specification has been considered. This test should follow the logical specification of whether the μ_i are truly random. Thus, the situation is very similar to simultaneous equation estimation which follows the logical question of identification. In the next section, the specification of simultaneous equation systems is considered, and a test is developed for correct system specification.

5. Specification of Simultaneous Equation Systems

Most estimation associated with simultaneous equation models has used single equation, limited information estimators. Thus, two stage least squares (2SLS) is by far the most widely used estimator. If a simultaneous equation system is estimated equation by equation, no check on the "internal consistency" of the entire specification is made. An important potential source of information on misspecification is thus neglected. This neglect is not total; one class of tests compares estimates of the unrestricted reduced form model with the estimates of the structural model as a test of the overidentifying restrictions.[1.] Unfortunately, this type of test has not been been widely used. Perhaps the reason has been the inconvenience of calculating the likelihood value or the nonlinear expansions which are required to perform the statistical comparison. In this section a test of system specification is proposed within a more simple framework. The test rests on a comparison of 2SLS to 3SLS estimates. Under the null hypothesis of correct specification, 3SLS is efficient but yields inconsistent estimates of all equations if any equation is misspecified. 2SLS is not as efficient as 3SLS, but only the incorrectly specified equation is inconsistently estimated if misspecification is present in the system.

Consider the standard linear simultaneous equation model

$$(5.1)\quad YB + Z\Gamma = U$$

---

1. Within the single equation context this test has been proposed by Anderson and Rubin [1949], Basmann [1957], and Hood and Koopmans [1953]. Within the full information context the likelihood ratio (LR) test has been used. Recently, Byron [1972, 1974] has simplified this test by advocating use of the Lagrange multiplier test or the Wald test both of which are asymptotically equivalent to the LR test under the null hypothesis. For further details see Silvey [1970, Ch. 7].

where Y is the $T \times M$ matrix of jointly dependent variables, Z is the $T \times K$ matrix of predetermined variables, and U is a $T \times M$ matrix of structural disturbances of the system. Full column rank of Z, nonsingularity of B, nonsingular probability limits of second order moment matrices, and the rank condition for identification are all assumed to hold. The structural disturbances are multivariate normal $U \sim N(0, \Sigma \otimes I_T)$. After a choice of normalization and imposition of zero restrictions each equation is written

$$(5.2) \quad y_i = X_i\delta_i + U_i \text{ where } X_i = [Y_i \; Z_i] \text{ and } \delta_i = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix}$$

where β_i has r_i elements and γ_i has s_i elements which correspond to the variables in X_i whose coefficients are not known a priori to be zero. It is convenient to stack the M equations into a system

$$(5.3) \quad y = X\delta + U$$

$$\text{where} \quad y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_M \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \cdot \\ \cdot \\ \cdot \\ \delta_M \end{bmatrix}, \quad U = \begin{bmatrix} U_1 \\ \cdot \\ \cdot \\ \cdot \\ U_M \end{bmatrix}.$$

The two stage least squares estimator when used on each equation of the system can conveniently be written in stacked form as $\hat{\delta}_{2SLS} = (X'\tilde{P}_Z X)^{-1}X'\tilde{P}_Z y$ where $\tilde{P}_Z = I_M \otimes Z(Z'Z)^{-1}Z'$. To simplify notation rewrite the estimator as $\hat{\delta}_2 = (\hat{X}'\hat{X})^{-1}\hat{X}'y$. Three stage least squares uses full information and links together all equations of the system through the estimate of the covariance matrix $\hat{\Sigma}$. Letting $\tilde{P}_{\Sigma Z} = \hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z'$, the 3SLS estimator is $\hat{\delta}_{3SLS} = (X'\tilde{P}_{\Sigma Z}X)^{-1}X'\tilde{P}_{\Sigma Z}y$ which is simplified to

$\hat{\delta}_3 = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y.$[1] Now 3SLS transmits misspecification throughout the entire system, affecting the estimates of all coefficients since $\hat{\delta}_3 - \delta = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'U$. Thus, if the j^{th} equation is misspecified plim $\frac{1}{T}\hat{X}_j'U_j \neq 0$, and so assuming probability limits exist with $\tilde{\Sigma}$ being the probability limit of the inconsistent estimate of Σ with $\tilde{\sigma}^{ij}$ the element of its inverse, the inconsistency is calculated from plim $(\hat{\delta}_3 - \delta) = \text{plim}\,(\frac{1}{T}\tilde{X}'\tilde{X})^{-1}\cdot$ plim $(\frac{1}{T}\tilde{X}'U)$. Looking at the crucial last term more closely, consider the unknown elements from the first equation δ_1. The last term takes the form

$$(5.4)\quad \text{plim}\,(\frac{1}{T}\tilde{X}'U) = \text{plim}\,\frac{1}{T}\sum_{m=1}^{M}\tilde{\sigma}^{1m}\hat{X}_m'U_m$$

so that the amount of inconsistency for the first equation due to misspecification in the j^{th} equation depends both on the lack of orthogonality between $\hat{X}_j$ and U_j, and also on the size of $\tilde{\sigma}^{1j}$.

The lemma (2.1) leads us to consider the specification test based on the difference between the two estimators $\hat{q} = \hat{\delta}_2 - \hat{\delta}_3$ which has large sample variance $V(\hat{q}) = V(\hat{\delta}_2) - V(\hat{\delta}_3)$. However, a simpler procedure is to consider the regression on the stacked system

$$(5.5)\quad y = \hat{X}\tilde{\delta} + \tilde{X}\alpha + V$$

and to test if $\alpha = 0$. Since $\hat{X}$ and $\tilde{X}$ are computed by programs which have 2SLS and 3SLS estimators, the regression of equation (5.5) should not be difficult to perform. A χ^2 test, rather than an F test, is appropriate given the large sample nature of the estimators.

---

1. If $T \leq K$ so 2SLS and 3SLS cannot be used, asymptotically equivalent instrumental variable estimators are discussed in Brundy and Jorgenson [1971], Dhrymes [1971], and Hausman [1975].

The noncentrality parameter of the χ^2 distribution will be proportional to plim $\frac{1}{T}\hat{X}'U_j$ for any equation which is misspecified and also the magnitude of the covariance elements $\tilde{\sigma}^{ij}$. If the inverse covariance elements are large, then $\hat{X}$ and $\tilde{X}$ will not be highly correlated so that the test will be powerful for a given size of inconsistency. As the $\tilde{\sigma}^{ij}$'s go to zero, then 3SLS approaches 2SLS and the test will have little power. Since the misspecification represented by the alternative hypothesis is not specific, the appropriate action to take in the case of rejection of H_0 is not clear. One only knows that misspecification is present somewhere in the system. If one were confident that one or more equations are correctly specified, then the specification of other equations could be checked by using them, say one at a time, to form a 3SLS type estimator. That is, if equation 1 is correct and equation 2 is to be tested, then 2SLS on equation 1 could be compared to 3SLS on equation 1 where $\hat{\sigma}_{ij}$ is set to zero for $i \neq j$ except for $i = 1$, $j = 2$ and vice-versa in the 3SLS estimator. Using this method the misspecification might be isolated; but, unfortunately, the size of the test is too complicated to calculate when done on a sequence of equations.

The simultaneous equations specification test is the last to be presented although the same principle may be applied to further cases. Before presenting two examples of the test, the issue of minimum mean square estimators (MMSE) is discussed. These estimators might be thought to be appropriate to use if the null hypothesis is rejected when the condition of unbiasedness is relaxed.

6. Pretesting and Minimum Mean Square Error Estimation

All the specification tests discussed so far have a single purpose: to test whether the specification of the statistical model is correct. The sample at hand is used for this purpose; and with respect to formal received theory, that should be the end of the story. However, upon deciding whether or not to reject H_0, the same data is often used to attempt further inferences utilizing the estimator which the specification test indicated is "correct". For example, the regression specification $y = X\beta + \varepsilon$ may have an associated test of the hypothesis H_0: $R\beta = r$ (Theil's specification theorem concerns a subvector of β being zero). After using an F test to determine whether to reject H_0, either the restricted or unrestricted least squares estimator is used to provide estimates for further inference. The properties of these so-called "pretest" estimators were first studied in a classic paper by Bancroft [1944] who showed that both bias and loss of efficiency may be introduced using such procedures. A long list of papers on pretest estimators has followed which will not be reviewed here. The point of this discussion is to mention the fact that only if the restrictions are imposed a priori, presumably because they are known to be true without any pretesting, do the classical statistical properties hold.

Besides the problem of pretesting, the other issue common to specification tests is minimum mean square error (MMSE) estimators. The mean square error (MSE) of an estimator is the bias squared plus the variance. The classical estimators of econometrics are all limited to unbiased estimators which then minimize the variance under appropriate statistical assumptions. In fact, one interpretation of the specification tests proposed here is to determine if the models satisfy these statistical

assumptions. Considered as an optimization problem, the MSE can be decreased by permitting some bias. Within the context of the models considered here, even if H_0 is rejected the size of the test determines the probability that H_0 is actually true. Thus, a weighted average of the two estimators will do better in a MSE context at least part of the time. Presumably, the relative weight given to the efficient estimator under H_0 is large when H_0 is thought not to be incorrect. The relative weight would then decrease as the evidence favoring rejection of H_0 increases. These weighted average estimators were first introduced by Huntsberger [1955] and have been suggested recently by Feldstein [1973, 1974] and Mundlak [1976]. Now a classical objection to MMSE type estimators in general is that choosing as our estimator $\hat{\beta} = (\pi,\ldots,\pi)$ will be unbeatable if all the elements of the true β vector are indeed π. This type example is the basis for proofs that in general MMSE estimators do not exist. However, an additional restriction that the new estimator must be a weighted average of the estimators appropriate to H_0 and H_1, $\hat{\beta}_* = \hat{\lambda}\hat{\beta}_0 + (1-\hat{\lambda})\hat{\beta}_1$, might be thought to provide sufficient limitations to solve the problem. For example, a simple case considered by Bancroft, Huntsberger and Feldstein is the trivariate regression specification

$$(6.1) \quad y = \beta x + \gamma z + \varepsilon$$

where $\hat{\beta}_0$ is from the bivariate regression and $\hat{\beta}_1$ from the full trivariate regression. The optimal value of λ,

$$(6.2) \quad \lambda^* = \frac{\gamma^2 D}{(x'x)\sigma^2 + \gamma^2 D} \text{ where } D = \det\,[(x'x)(z'z) - (x'z)^2]$$

contains unknown parameters so $\hat{\lambda}$ uses estimates derived from $\hat{\beta}_1$ and s_1^2. Since the weight $\hat{\lambda}$ will depend on the test statistic of H_0 versus

H_1 which is consistent, $\hat{\lambda}$ will go to zero as the sample becomes large if H_0 is not true. Thus, in large samples the estimate is consistent but the MSE properties of interest will usually be finite sample properties since for large samples the pretest estimator ($\hat{\lambda} = 0$ or $\hat{\lambda} = 1$) will lead to the correct estimator. For finite samples, even within the restricted class of weighted average estimators no MMSE estimator will exist in general. This result can be seen easily since $\hat{\lambda}$ is a statistic and for H_0 being nearly true the weighted average estimator will not give enough weight to $\hat{\beta}_0$ due to the uncertainty of the true λ. Statements which attempt to give guidance about which estimator to use (i.e., a priori choice of $\hat{\lambda}$) conditional on the researcher's "confidence" about the size of unknown parameters or unknown test statistics seem an undesirable form of "back-door Bayesianism". The correct use of prior knowledge is provided by Bayesian analysis which is superior to such "rules of thumb" estimators.[1.]

The other rationale sometimes advocated in the use of such estimators is that the researcher is not interested in certain of the parameters. With unbiased estimators, Gauss-Markov or minimum variance estimation assures us that $c'\hat{\beta}$ is the best estimator of $c'\beta$ for $\hat{\beta}$ minimum variance and c an arbitrary vector. For certain choices of c within the MSE framework, it is true that unbiased minimum variance estimators are no longer uniformly best. But, neither will MMSE estimator be best in all situations just as before. Furthermore, since other researchers are unlikely to have the same c, the correct procedure is to report both $\hat{\beta}_0$ and

---

1. It is interesting to note that for any unknown β, the unbiased OLS estimator may be uniformly improved upon for any quadratic loss function if the unknown t statistic for γ is known to be bounded. While Feldstein [1973] advocates use of the t statistic as a rule of thumb, his estimator is not the one which offers this uniform improvement. For details see Perlman [1972].

$\hat{\beta}_1$ along with their associated covariance matrices. Users of the research can then decide if they are convinced by the tests conducted and apply their own weights to form their favorite estimates. Reporting only the weighted estimates is condensing the original data too far since the results depend on either the original researcher's confidence about λ or choice of c which are unlikely to be shared by his readers.

7. Empirical Examples

Comparing two alternative estimators as a means of constructing misspecification tests has been applied to a number of situations in the preceding sections. In this section two empirical examples are presented. Both examples are new tests of misspecification and likely to be of interest to the applied econometrician. The first example is the time series - cross section specification test discussion in Section 4. This type of data set is becoming increasingly common for econometric studies such as individuals' earnings, education, and labor supply. However, added interest in this test comes from the fact that it also implicitly tests much cross section analysis of similar specifications. Cross section analysis can allow for no individual constant but must assume, as does random effect analysis, that the right hand side variables are orthogonal to the residual: If the random effect specification is rejected serious doubt may be cast therefore on much similar cross section analysis. The second empirical example is the simultaneous equations specification test of Section 5. The famous Klein Model I is used as an example since it has been thoroughly analyzed in the past and is a convenient example. Previous tests of the model have been tests mainly of the overidentifying restrictions on the structural form. Here by comparing 2SLS and 3SLS estimates of the model, the correctness of the overall specification is tested.

For the time series - cross section specification test a wage equation is estimated for male high school graduates in the Michigan Income Dynamics Study.[1] The sample consists of 629 individuals for whom all six years

1. The specification used is based on research by Gordon [1976] who also kindly helped me construct this example.

of observations are present. A wage equation has been chosen due to its importance in "human capital" analysis. The specification used follows from equation (4.1). The right hand side variables include a piecewise linear representation of age, the presence of unemployment or poor health in the previous year, and dummy variables for self-employment, living in the South or in a rural area. The fixed effects estimates, $\hat{\beta}_{FE}$, are calculated from equation (4.3). They include an individual constant for each person and are unbiased under both the null hypothesis of no misspecification and the alternative hypothesis. The random effects estimates, $\hat{\beta}_{GLS}$, are calculated from equations (4.4)-(4.6). The estimate of $\hat{\gamma}$ from equation (4.6) is .72736 which follows from least squares estimates of the individual variance $\hat{\sigma}_{\mu}^2 = .12594$ and the residual variance $\hat{\sigma}_{\varepsilon}^2 = .06068$. Under the null hypothesis the GLS estimate is asymptotically efficient, but under the alternative hypothesis it is inconsistent. The specification test consists of seeing how large the difference in estimates is, $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}$, in relation to its variance $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{GLS})$ which follows from Lemma (2.1). In comparing the estimates in column 1 and column 2 of Table 1 it is apparent that substantial differences are present in the two sets of estimates relative to their standard errors which are presented in column 3.[1] The effects of unemployment, self-employment, and geographical location differ widely in the two models. For instance, the effect of unemployment in the previous year is seen

---

1. Note that the elements of $\hat{q}$ and its standard errors are simply calculated given the estimates of $\hat{\beta}_{FE}$ and $\hat{\beta}_{GLS}$ and their standard errors making sure to adjust to use the fixed effects estimate of σ_{ε}^2. The main computational burden involves forming and inverting $V(\hat{q})$.

Table I: Dependent Variable - Log Wage

observations = 3774, (standard error)

	Var	Fixed Effects	Random Effects	$\hat{q}$	$\hat{\alpha}$
1.	Age 1 (20-35)	.0557	.0393	.0164	.0291
		(.0042)	(.0033)	(.0030)	(.0060)
2.	Age 2 (35-45)	.0351	.0092	.0259	.0015
		(.0051)	(.0036)	(.0039)	(.0070)
3.	Age 3 (45-55)	.0209	-.0007	.0216	.0058
		(.0055)	(.0042)	(.0040)	(.0083)
4.	Age 4 (55-65)	.0209	-.0097	.0306	-.0308
		(.0078)	(.0060)	(.0050)	(.0112)
5.	Age 5 (65-)	-.0171	-.0423	.0252	-.0380
		(.0155)	(.0121)	(.0110)	(.0199)
6.	Unemployed_{-1}	-.0042	-.0277	.0235	-.3290
		(.0153)	(.0151)	(.0069)	(.0914)
7.	Poor Health_{-1}	-.0204	-.0250	.0046	-.1716
		(.0221)	(.0215)	(.0105)	(.0762)
8.	Self-Employment	-.2190	-.2670	.0480	-.3110
		(.0297)	(.0263)	(.0178)	(.0558)
9.	South	-.1569	-.0324	-.1245	.0001
		(.0656)	(.0333)	(.0583)	(.0382)
10.	Rural	-.0101	-.1215	.1114	-.2531
		(.0317)	(.0237)	(.0234)	(.0352)
11.	Constant	-	.8499	-	
		-	(.0433)	-	
	s^2	.0567	.0694		.0669
	degrees of freedom	3135	3763		3753

to be much less important in effecting the wage in the fixed effects specification. Thus, unemployment has a more limited and transitory effect once permanent individual differences are accounted for. The test of misspecification follows from Lemma 2.1 is

$$(7.1) \quad m = \frac{\hat{q}' V(\hat{q})^{-1} \hat{q}}{10} = 12.99.$$

Since m is distributed approximately as $F(10, \infty)$ which has a critical value of 2.32 at the 1% level, very strong evidence of misspecification in the random effects model is present. The right hand side variables X_{it} are not orthogonal to the individual constant μ_i so that the null hypothesis is decisively rejected. Considerable doubt about much previous cross section work on wage equations arises from this example.[1.]

The reason for this doubt about previous cross section estimation is that ordinary least squares on a cross section of one year will have the same expectation as $\hat{\beta}_{GLS}$, the random effects estimate, on the time series - cross section data. For example, cross section estimates of the wage equation have no individual constants and make Assumption (1.1.a.) that the residual is uncorrelated with the right hand side variables. However, this example demonstrates that in the Michigan Survey important individual effects are present which are <u>not</u> uncorrelated with the right hand variables. Since the random effects estimates seem significantly biased with high probability, then previous cross section estimates of wage and earnings equations may also be significantly biased. This problem can only be resolved within a time series - cross section frame-

---

1. Direct estimates of the effect of education are not possible in the fixed effects approach, but the example shows that models which use this specification may well be misspecified.

work using a specification which allows testing of an important maintained hypothesis of much cross section estimation in econometrics.

An equivalent formulation of the specification test is provided by the regression framework of equation (4.7). Instead of having to manipulate 10 × 10 matrices, $\check{y}$ is regressed on both $\check{X}$ and $\tilde{X}$. The test of the null hypothesis is then whether $\hat{\alpha} = 0$. As is apparent from column 4 of Table 1 many of the elements of $\hat{\alpha}$ are well over twice their standard error so that misspecification is clearly present. The misspecification test follows easily from comparing s^2, the estimated variance, from the random effects specification to s^2 from the augmented specification

$$(7.2)\quad m = \frac{.06938 - .06689}{.06689} \cdot \frac{3754}{10} = 13.974.$$

Again m well exceeds the approximate critical F value of 2.32. Since this form of the test is so easy to implement when using a random effects specification as only one additional weighted least squares regression is required, hopefully applied econometricians will find it a useful device for testing specification.

The second empirical example is a test of Klein Model I. This seminal model has 3 equations for consumption, investment, and labor and is estimated on annual data from 1920-1941. It is known that the hypothesis of the overidentifying restrictions is rejected for the model. Thus, the misspecification test may not have great power since under just identification the 2SLS and 3SLS estimates are identical. Still, the test may allow us to derive further evidence about the model. Here 2SLS estimates will be consistent for all but the misspecified equation under the alternative

specification while the 3SLS estimates for all equations will be inconsistent. Another determinant of the power of the test is the covariance matrix Σ since if it is diagonal 2SLS and 3SLS estimates are again identical. The 2SLS estimate $\hat{\Sigma}$, however, shows substantial covariance between the equations.

$$(7.3) \quad \hat{\Sigma}_{2SLS} = \begin{bmatrix} 1.044 & & \\ .4378 & 1.383 & \\ -.3852 & .1926 & .4764 \end{bmatrix}$$

When comparing the 2SLS and 3SLS estimates in Table 2, the estimated coefficients are quite similar relative to their standard errors as seen in column 3. Thus $\hat{q} = \hat{\delta}_2 - \hat{\delta}_3$ does not present much evidence of misspecification. Forming a χ^2 test from $\hat{q}$ and its estimated variance $V(\hat{q}) = V(\hat{\delta}_2) - V(\hat{\delta}_3)$ leads to a value

$$(7.4) \quad m = \hat{q}'V(\hat{q})^{-1}\hat{q} = 12.71$$

Since m is distributed as χ^2_{12}, the test presents little evidence in favor of misspecification since the expected value of m under the null hypothesis is 12. Whenever the null hypothesis is not rejected, the power of the test is of considerable interest. Here power considerations are evident in using the stacked regression formulation of equation (5.5) to check the estimates of   relative to their standard errors. With this alternative approach m is calculated to be 5.78 which is distributed as χ^2_7 so no evidence of misspecification is present.[1] However, by considering the standard errors the test is seen not to have great power since $\hat{\alpha}$ from equation

1. In the combined regression framework of equation (5.5) constants are eliminated from $\tilde{X}$ and each right hand side variable appears only once. Therefore, in the stacked framework, the operating characteristics of the alternative tests are not identical.

Table II: Klein Model I

observations = 21, (standard error)

	2SLS	3SLS	$\hat{q}$	$\hat{\alpha}$
I. Consumption				
1. Constant	16.55	16.44	.11	-
	(1.468)	(1.305)	(.672)	
2. Profits	.0173	.1249	-.1076	-.0518
	(.1312)	(.1081)	(.0743)	(.3833)
3. Profits_{-1}	.2162	.1631	.0531	.1111
	(.1192)	(.1004)	(.0643)	(.5017)
4. Wage	.8102	.7901	.0201	-.0173
	(.0447)	(.0379)	(.0237)	(.1531)
II. Investment				
1. Constant	20.28	28.18	-7.90	-
	(8.383)	(6.794)	(4.911)	
2. Profits	.1502	-.0131	.1633	-
	(.1925)	(.1619)	(.1041)	
3. Profits_{-1}	.6159	.7557	-.1398	-
	(.1809)	(.1529)	(.0967)	
4. Capital_{-1}	-.1578	-.1948	.0370	-.0294
	(.0401)	(.0325)	(.0235)	(.1488)
III. Labor				
1. Constant	1.500	1.797	-.297	-
	(1.276)	(1.116)	(.619)	
2. Production	.4389	.4005	.0384	-.0381
	(.0396)	(.0318)	(.0236)	(.2751)
3. Production_{-1}	.1467	.1813	-.0346	.0395
	(.0432)	(.0342)	(.0264)	(.6956)
4. Time_{-1931}	.1304	.1497	-.0193	.0781
	(.0323)	(.0279)	(.0163)	(.2220)

(5.5) is not at all precisely estimated. Some of the elements of $\hat{\alpha}$ are large relative to their estimated value in $\hat{\delta}$ from equation (5.5), e.g. profits, but the estimated standard errors are so large that the test cannot determine if this result follows from misspecification or from statistical fluctuation.

The two empirical examples presented in this section illustrate use of the misspecification test. The first example rejects an application of the random effects specification and thereby casts doubt on much cross section work in this area. I feel that this finding is probably quite general, and that the random effects model is not well suited to most econometric applications. The two requirements of exchangeability and orthogonality are not likely to be met in our applied problems. Certainly, these specifications should be tested for correct specification. The second example demonstrates how power considerations are evident when the null hypothesis is not rejected. Also, it demonstrates the potential usefulness of full information estimators in determining the correctness of specification in simultaneous equation models.

8. Conclusion

By using the result that under the null hypothesis of no misspecification, an efficient estimator must be uncorrelated with its difference from an unbiased but inefficient estimator, specification tests are devised for a number of important model specifications in econometrics. New tests for the cross section - time series model and for the simultaneous equation model are presented. The possibility of combining the two estimators into a MMSE estimator is discussed, and it is pointed out that the type of knowledge needed for such estimators is better used within a proper Bayesian framework. Lastly, two empirical examples are provided. The first example provides strong evidence against a specification commonly used in time series - cross section work and also provides evidence questioning much cross section analysis currently being done on individual data in econometrics.

References

Anderson, T.W. and Rubin, H., "Estimation of the Parameters of a Single Stochastic Equation in a Complete System of Stochastic Equations", Annals of Mathematical Statistics, 20, 1949, pp. 46-63

Bancroft, T.A., "On Biases in Estimation due to the Use of Preliminary Tests of Significance", Annals of Mathematical Statistics, 15, 1944, pp. 190-204

Basmann, R.L., "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation", Econometrica, 25, 1957, pp. 77-83

Brundy, J. and Jorgenson, D.W., "Efficient Estimation of Simultaneous Equation Systems by Instrumental Variables, Review of Economics and Statistics, 53, 1971, pp. 207-224

Byron, R.P., "Testing for Misspecification in Econometric Systems Using Full Information", International Economic Review, 13, 1972, pp. 745-756

________________, "Testing Structural Specification Using the Unrestricted Reduced Form", Econometrica, 42, 1974, pp. 869-883

Dhrymes, P.J., "A Simplified Structural Estimator for Large-Scale Econometric Models", Australian Journal of Statistics, 13, 1971, pp. 168-175

Feldstein, M., "Multicollinearity and the Mean Square Error of Alternative Estimators", Econometrica, 41, 1973, pp. 337-346

________________, "Errors in Variables: A Consistent Estimator with Smaller Mean Square Error in Finite Samples", Journal of American Statistical Association, 69, 1974, pp. 990-996

Gordon, R., "Essays on the Causes and Equitable Treatment of Differences in Earnings and Ability", MIT PhD Thesis, June 1976

Griliches, Z., "Specification Bias in Estimates of Production Functions", Journal of Farm Economics, 39, 1957, pp. 8-20

Hausman, J., "An Instrumental Variable Approach to Full-Information Estimators for Linear and Certain Nonlinear Econometric Models", Econometrica, 43, 1975, pp. 727-738

Huntsberger, D.V., "A Generalization of a Preliminary Test Procedure for Pooling Data, Annals of Mathematical Statistics, 26, 1955, pp. 734-743

Koopmans, T.C. and Hood, W., "The Estimation of Simultaneous Economic Relationships", in Hood and Koopmans (eds.), Studies in Econometric Method, New Haven, 1953, pp. 113-199

Liviatan, N., "Tests of the Permanent Income Hypothesis Based on a Re-interview Savings Study", in Measurement in Economics, C. Christ, ed., Stanford, 1963, pp. 29-59

Maddala, G.S., "The Use of Variance Components Models in Pooling Cross Section and Time Series Data", Econometrica, 39, 1971, pp. 341-358

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data", 1976, mimeo.

Nerlove, M., "A Note on Error Component Models", Econometrica, 39, 1971, pp. 383-396

Perlman, M.D., "Reduced Mean Square Error Estimation for Several Parameters", Sankhya, B, 34, 1972, pp. 89-93

Rao, C.R., Linear Statistical Inference, New York, 1973

Sargan, J.D., "The Estimation of Economic Relationships Using Instrumental Variables", Econometrica, 26, 1958, pp. 393-415

Scheffé, H., Analysis of Variance, New York, 1959

Searle, P., Linear Models, New York, 1971

Silvey, S.D., Statistical Inference, London, 1970

Theil, H., "Specification Errors and the Estimation of Economic Relationships", Review of the International Statistical Institute, 25, 1957, pp. 41-51

Wallace, T.D. and Hussain, A., "The Use of Error Components Models in Combining Cross Section with Time Series Data", Econometrica, 37, 1969, pp. 57-72

Wu, D., "Alternative Tests of Independence Between Stochastic Regressors and Disturbances", Econometrica, 41, 1973, pp. 733-750